Proposed Content Standard for Metadata for Tools for Biological Analysis

Based on:

Content Standards for Digital Geospatial Metadata

Federal Geographic Data Committee

June 8, 1994

and

DRAFT Content Standard for National Biological Information Infrastructure Metadata

National Biological Service
December 1995
Overview

1. Name of Standard. Content Standards for Metadata for Tools for Biological Analysis.

2. Explanation and Purpose. This standard specifies the information content of metadata for tools for performing biological analysis. Tools are *algorithms or software that help with the analysis, interpretation, application, or understanding of biological data.* Examples of tools for biological analysis include ecological models, decisionsupport systems, simulations, and visualizations.

This standard establishes the names of elements for metadata for tools, the definitions of these elements, and information about the values that are to be provided for the elements. The purpose of the standard is to provide a common set of terminology and definitions for describing metadata for tools. The standard is patterned after the "Content Standards for Digital Geospatial Metadata" of the US Federal Geographic Data Committee {http://www.fgdc.gov/Metadata/metav1-0.html} and the complementary Metadata Standard for the National Biological Information Infrastructure {http://www.nbii.gov/factsheet/factsheet3.html}.

The standard was developed from the perspective of defining the information required by a prospective user to determine the availability of tools for biological analysis, to determine the appropriateness of the tools for an intended use, and to determine how to access and use the tools. The standard establishes the names of data elements to be used for these purposes, the definitions of the data elements, and information about the values that are to be provided for the data elements.

3. Approving Authority.

4. Maintenance Authority.

5. Related Documents.

6. Objectives. The objectives of this standard are to provide a common set of terminology and definitions for the documentation of tools for performing biological analysis. There is no standard that currently serves this function, and this document is intended to serve this need. Metadata on *tools*, like metadata on *data,* are intended to help people discover and effectively apply tools for biological analysis.

7. Applicability. This standard is for the documentation of *tools for biological analysis*. It is intended for use by anyone who wishes to describe and document tools for biological analysis. The standard is intended to help potential tool users discover and apply tools to their tasks. The standard is intended to apply to any and all tools for performing biological analyses, such as ecological, physiological, and biomolecular models, simulations, landscape analyses, and

visualizations, at all levels of resolution and granularity.

8. Specifications. The standard provides definitions and specifications for data elements for describing properties of tools for biological analysis. Some elements are mandatory, others conditionally mandatory, and still others optional (i.e., supplied at the discretion of the information provider).

Table of Contents

Organization of the Standard

Numbered Sections

This standard is organized in a hierarchy of data elements and compound elements that define the information content for metadata to document a tool for biological analysis. The starting point is "metadata" (section 0). The compound element "metadata" is composed of other compound elements representing different concepts about the tool for biological analysis. Each of these compound elements has a numbered section in the standard. In each numbered section, these compound elements are defined by other compound elements and data elements. The sections "citation information", and "contact information" are special sections that specify the recommended reference to be used for a data set, and the data elements for contacting individuals and organizations respectively. These sections are used by other sections and are defined once for convenience.

Each section begins with the name and definition of the compound element that defines the section. The name and definition are followed by production rules (see below) that define this compound element in terms of data elements, either directly or by the use of intermediate compound elements. When intermediate compound elements are used, the production rules for these elements also are provided in this part of the section.

The production rules are followed by a list of names and definitions of compound elements and data elements used in the section.

Compound Elements

A compound element is a group of data elements and other compound elements. All compound elements are described by data elements, either directly or through intermediate compound elements. Compound elements represent higher-level concepts that cannot be represented by individual data elements. The form for the definition of compound elements is:

Compound element name -- definition.
Type: compound

The type of "compound" uniquely identifies the compound elements in the lists of terms and definitions.

Production Rules

A production rule specifies the relationship between a compound element and data elements and other (lower-level) compound elements. Each production rule has a left side (identifier) and a right side (expression) connected by the symbol "=", meaning that the term on the left side is replaced by or produces the term on the right side. Terms on the right side are either other compound elements or individual data elements. By making substitutions using matching terms in the production rules, one can explain higher-level concepts using data

elements.
The symbols used in the production rules have the following meaning:

Symbol	Meaning
=	is replaced by, produces, consists of
+	and
[|]	selection - select one term from the list of enclosed terms (exclusive or). Terms are separated by "|".
m{}n	iteration - the term(s) enclosed is(are) repeated from "m" to "n" times
()	optional - the term(s) enclosed is(are) optional

Examples:

a = b + c	**"a consists of b and c"**
a = [b | c]	**"a consists of one of b or c"**
a = 4{b}6	**"a consists of four to six occurrences of b"**
a = b + (c)	**"a consists of b and optionally c"**

Interpreting the production rules:

- **The terms bounded by parentheses, "(" and ")", are optional and are provided at the discretion of the data producer. If a producer chooses to provide information enclosed by parentheses, the producer shall follow the production rules for the enclosed information. For example, if the producer decides to provide the optional information described in the term:**

  (a + b + c)

  the producer shall provide a and b and c.

  Only for terms bounded by parentheses does the producer have the discretion of deciding whether or not to provide the information.

- **The variation among the ways in which data are produced and distributed, the fact that all data do not have the same characteristics, and the issue that all details of data sets that are in work or are planned may not be decided, makes it necessary to express the concept of "mandatory if applicable." This concept means that if the data set exhibits (or, for data sets that are in work or planned, it is known that the data set will exhibit) a defined characteristic, then the producer shall provide the information needed to describe that characteristic. This concept is described by the production rule:**

0{ term }1

Data Elements

A data element is a logically primitive item of data. The entry for a data element includes the name of the data element, the definition of the data element, and a description of the values that can be assigned to the data element. The form for the definition of the data element is:

Data element name -- definition.
Type:
Domain:

The information about the values for the data elements include a description of the type of the value, and a description of the domain of the valid values. The type of the data element describes the kind of value to be provided. The choices are "integer" for integer numbers, "real" for real numbers, "text" for ASCII characters, "date" for day of the year, and "time" for time of the day.

The domain describes valid values that can be assigned to the data element. The domain may specify a list of valid values, references to lists of valid values, or restrictions on the range of values that can be assigned to a data element.

The domain also may note that it is free from restrictions, and any values that can be represented by the type of the data element can be assigned. These unrestricted domains are represented by the use of the word "free" followed by the type of the data element (that is, free text, free date, free real, free time, free integer).

Some domains can be partly, but not completely, specified. For example, there are several widely used data transfer formats, but there may be many more that are less well known. To allow a producer to describe its data in these circumstances, the convention of providing a list of values followed by the designation of a "free" domain is used. In these cases, assignments of values shall be made from the provided domain when possible. When not possible, providers may create and assign their own value. A created value shall not redefine a value provided by the standard.

Another issue is the representation of null values (representing such concepts as "unknown") in the domain. While this is relatively simple for textual entries (one would enter the text "Unknown"), it is not as simple for the integer, real, date, and time types (for example, which integer value means "unknown"?). Because conventions for providing this information vary among implementations, the standard specifies what concepts shall be represented, but does not mandate a means for representing them.

In addition to the values to be represented, the form of the representation also is important, especially to applications that will manipulate the data elements. The following conventions for forms of values for data elements shall be used:

Calendar Dates (Years, Months, and Days)

- **A.D. Era to December 31, 9999 A.D. -- Values for day and month of year and for years shall follow the calendar date convention (general forms of YYYY for years; YYYYMM for month of a year (with month being expressed as an integer), and YYYYMMDD for a day of the year) specified in American National Standards Institute, 1986, Representation for Calendar Date and Ordinal Date for Information Interchange (ANSI X3.30-1985): New York, American National Standards Institute (adopted as Federal Information Processing Standard 4-1).**

- **B.C. Era to 9999 B.C. -- Values for day and month of year, and for years, shall follow the calendar date convention, preceded by the lower case letters "bc" (general forms of bcYYYY for years; bcYYYYMM for month of a year (with month being expressed as an integer), and bcYYYYMMDD for a day of the year).**

- **B.C. Era before 9999 B.C. -- Values for the year shall consist of as many numeric characters as are needed to represent the number of the year B.C., preceded by the lower case letters "cc" (general form of ccYYYYYYY...).**

- **A.D. Era after 9999 A.D. -- Values for the year shall consist of as many numeric characters as are needed to represent the number of the year A.D., preceded by the lower case letters "cd" (general form of cdYYYYYYY...).**

Network Addresses and File Names

- **values for file names, network addresses for computer systems, and related services should follow the Uniform Resource Locator convention of the Internet when possible. See http://www.ncsa.uiuc.edu/demoweb/url-primer.html for additional details about the Uniform Resource Locator.**

Metadata

0 Metadata -- data about the identification, algorithms used, source of, and other characteristics of tools for biological analysis.
Type: compound

Metadata =
Identification_Information +
0{Quality_Information}1 +
0{Distribution_Information}1 +
Metadata_Reference_Information

(Sections 1 through 4 define the terms on the right side of the production rule.)

Identification_Information
= the basic information about the tool for biological analysis.

Quality_Information
= a general assessment of the quality of the biological analysis tool.

Distribution_Information
= information about the distributor of and options for obtaining the tool for biological analysis.

Metadata_Reference_Information
= information on the currentness of the metadata information and the responsible party.

Identification Information

1 Identification Information -- basic information about the tool for biological analysis.
Type: compound

Identification_Information =
Citation +
Description +
Status +
1{Keywords}n +
Access_Constraints +
Use_Constraints +
(Point_of_Contact) +
(1{Browse_Graphic}n) +
(Credit) +
(Security_Information) +
(Native_Computer_Environment) +
(1{Associated_Work}n)

Citation =

Citation_Information ***(see section 5 for production rules)***

Description =

Abstract +
Purpose +
Language +
(Known_Uses) +
(Supplemental_Information)

Status =

Progress +
Maintenance_and_Update_Frequency

Keywords = **Keyword_Type +**
Keyword_Thesaurus +
Keyword

Point_of_Contact =

Contact_Information ***(see section 6 for production rules)***

Browse_Graphic =

Browse_Graphic_File_Name +
Browse_Graphic_File_Description +
Browse_Graphic_File_Type

Security_Information =

Security_Classification_System +
Security_Classification +
Security_Handling_Description

Associated_Work =

Reference_Type +
Citation_Information ***(see section 5 for production rules)***

1.1 **Citation -- information to be used to reference the tool for biological analysis.**
Type: compound

1.2 **Description -- a characterization of the tool for biological analysis, including its intended use and limitations.**
Type: compound

1.2.1 **Abstract -- a brief narrative summary of the tool for biological analysis.**
Type: text
Domain: free text

1.2.2 **Purpose -- a summary of the purpose of the tool for biological analysis.**
Type: text
Domain: free text

1.2.3 **Language -- language(s) used within the tool for biological analysis.**
Type: text
Domain: free text

1.2.4 **Known Uses -- brief description of the known uses of the tool for biological analysis, if different from 1.2.2 Purpose.**
Type: text
Domain: free text

1.2.5 **Supplemental Information -- other descriptive information about the tool for biological analysis.**
Type: text
Domain: free text

1.3 **Status -- the status of and maintenance information for the tool for biological analysis.**
Type: compound

1.3.1 **Progress -- the status of the tool for biological analysis.**
Type: text
Domain: "Complete" "In work" "Planned"

1.3.2 **Maintenance and Update Frequency -- the frequency with which changes and additions are made to the *tool for biological analysis* after the initial *tool for biological analysis* is completed.**

Type: text

Domain: "Continually" "Daily" "Weekly" "Monthly" "Annually" "Unknown" "As needed" "Irregular" "None planned" free text

1.4 **Keywords -- words or phrases summarizing an aspect of the tool for biological analysis.**

Type: compound

1.4.1 ***Keyword Type -- method used to group similar keywords.***

Type: text

Domain: "theme" "temporal" "discipline" "taxonomic"

1.4.2 **Keyword Thesaurus -- reference to a formally registered thesaurus or a similar authoritative source of keywords. *For a list of some commonly-used thesauri, see Part IV: Subject/index term sources in Network Development and MARC Standards Office, 1988, USMARC code list for relators, sources, and description conventions: Washington, Library of Congress.***

Type: text

Domain: "None" free text

1.4.3 **Keyword -- common-use word or phrase used to describe the *tool for biological analysis*.**

Type: text

Domain: free text

1.5 **Access Constraints -- restrictions and legal prerequisites for accessing the *tool for biological analysis*. These include any access constraints applied to assure the protection of privacy or intellectual property, and any special restrictions or limitations on obtaining the *tool for biological analysis*.**

Type: text

Domain: "None" free text

1.6 **Use Constraints -- restrictions and legal prerequisites for using the *tool for biological analysis* after access is granted. These include any access constraints applied to assure the protection of privacy or intellectual property, and any special restrictions or limitations on obtaining the *tool for biological analysis*.**

Type: text

Domain: "None" free text

1.7 **Point of Contact -- contact information for an individual or organization that is knowledgeable about the *tool for biological analysis*.**

Type: compound

1.8 **Browse Graphic -- a graphic that provides an illustration *of any aspect* of the *tool for biological analysis*. The graphic should include a legend for interpreting the graphic. *This may include, but not be limited to, flow charts, diagrams, or pictures of equipment or procedures.***

Type: compound

1.8.1 **Browse Graphic File Name -- name of a related graphic file that provides an illustration of the *tool for biological analysis*.**

Type: text

Domain: free text

1.8.2 **Browse Graphic File Description -- a text description of the illustration.**

Type: text

Domain: free text

1.8.3 **Browse Graphic File Type -- graphic file type of a related graphic file.**

Type: text

Domain: domain values in the table below; free text

Domain Value	Definition
"CGM"	Computer Graphics Metafile
"EPS"	Encapsulated Postscript format
"GIF"	Graphic Interchange Format
"JPEG"	Joint Photographic Experts Group format
"PBM"	Portable Bit Map format
"PS"	Postscript format
"TIFF"	Tagged Image File Format
"XWD"	X-Windows Dump

1.9 **Credit -- recognition *or acknowledgment* of those who contributed to the *tool for biological analysis*.**

Type: text

Domain: free text

1.10 **Security Information -- handling restrictions imposed on the *tool for biological analysis* because of national security, privacy, or other concerns.**

Type: compound

1.10.1 **Security Classification System -- name of the classification system.**

Type: text
Domain: free text

1.10.2 **Security Classification -- name of the handling restrictions on the *tool for biological analysis*.**
Type: text
Domain: "Top secret" "Secret" "Confidential" "Restricted" "Unclassified" "Sensitive" free text

1.10.3 **Security Handling Description -- additional information about the restrictions on handling the *tool for biological analysis*.**
Type: text
Domain: free text

1.11 **Native *Computer* Environment -- a description of the *tool for biological analysis* producer's *production* environment, including items such as the name of the computer operating system.**
Type: text
Domain: free text

1.12 ***Associated Work* -- information about other, related *tool for biological analysiss and information products* that are likely to be of interest.**
Type: compound

1.12.1 ***Reference_Type -- type of reference being cited.***
Type: text
Domain: "related tool" "manual" "workbook" "related data set" "larger work" free text

Quality Information

2 **Quality Information -- a general assessment of the quality of the *tool for biological analysis. This field is applicable unless the tool for biological analysis is only planned and no information concerning tool quality has been proposed.***

Type: compound

Quality_Information =

Lineage +
Specification +
Compliance +
(Discussion)

Lineage =

0{Data_Set_Information}n +
1{Process_Step}n

Data_Set_Information =

Data_Set_Citation +
0{Data_Set_Form}1 +
(Data_Set_Citation_Abbreviation) +
Data_Set_Contribution

Data_Set_Citation =

Citation_Information (see section 5 for production rules)

Process_Step =

Process_Description +
0{Data_Set_Used_Citation_Abbreviation}n +
(*Process_Date*) +
0{Data_Set_Produced_Citation_Abbreviation}n +
(Process_Contact)

Process_Contact =

Contact_Information ***(see section 6 for production rules)***

2.1 **Lineage -- information about the *process steps*, parameters, and source data which *contribute to* the *tool for biological analysis*, and information about the responsible parties.**

Type: compound

2.1.1 ***Data Set Information -- description of data set used as input to a process step and a short discussion of the information contributed by each input data set.***

Type: compound

2.1.1.1 ***Data_Set Citation -- reference for the description of a data set which can be input to a process step of the tool for biological analysis.***

Type: compound

2.1.1.2 ***Data Set Form*** **-- the mode in which the *input data set* is represented.** ***This field is applicable if the data set is in a describable form.***

Type: text

Domain: (the domain is *based on* pp. 88-91 *in* Anglo-American Committee on Cataloguing of Cartographic Materials, 1982, Cartographic materials: A manual of interpretation for AACR2: Chicago, American Library Association) *with additions*:

"atlas" "diagram" "globe" "map" "model" "profile" "database" "graph" "remote-sensing image" "section" "view" ***"table" free text***

2.1.1.3 ***Data Set Citation Abbreviation -- short-form alias for the input data set citation.***

Type: text

Domain: free text

2.1.1.4 ***Data Set Contribution -- brief statement identifying the information contributed by the input data set to the biological analysis.***

Type: text

Domain: free text

2.1.2 **Process Step -- information about a single event.** ***Each processing step taken in the tool for biological analysis should be documented (e.g. data manipulation, statistical tests, algorithms, etc.).***

Type: compound

2.1.2.1 **Process Description -- an explanation of the event, *data manipulation, or algorithm* and related parameters or tolerances.**

Type: text

Domain: free text

2.1.2.2 ***Data Set Used Citation Abbreviation -- the Data Set Citation Abbreviation of a data set used in the processing step.***

Type: text

Domain: Data Set Citation Abbreviations from the Data Set Information entries for the tool for biological analysis.

2.1.2.3 **Process Date -- the date when the *process step* was *added to the tool for biological analysis*.**

Type: date

Domain: "Unknown" "Not complete" free date

2.1.2.4 ***Data Set Produced Citation Abbreviation -- the Data Set Citation Abbreviation of an intermediate data set that (1) is significant in the opinion of the data producer, (2) is generated by the processing step, and (3) is used in later processing steps.***

Type: text

Domain: Source Citation Abbreviations from the Source Information entries for the data set.

2.1.2.5 **Process Contact -- the party responsible for the process step.**

Type: compound

2.2 ***Specification -- what standard or specification it the tool based on, or developed under? Provide a citation if applicable.***

Type: text

Domain: "None" free text

2.3 ***Compliance -- does the tool comply with the standard, or meet the specification?***

Type: text

Domain: "Yes" "no" "Unknown"

2.4 ***Discussion -- any comments concerning the compliance of the tool with the standard or specification.***

Type: text

Domain: free text

Distribution Information

3 **Distribution Information -- information about the distributor of and options for obtaining the *tool for biological analysis*. *This field is applicable unless the tool for biological analysis is totally unavailable and there is no expected date when it will become available for distribution.***

Type: compound

Distribution_Information =
1{Distributor +
0{*Local_Identifier*}1 +
Distribution_Liability +
0{Standard_Order_Process}n +
0{Custom_Order_Process}1 +
(Technical_Prerequisites) +
(Available_Time_Period) }n

Distributor =
Contact_Information ***(see section 6 for production rules)***

Standard_Order_Process =
[Non-digital_Form |
1{Digital_Form}n] +
Fees +
(Ordering_Instructions) +
(Turnaround)

Digital_Form =
Digital_Transfer_Information +
Digital_Transfer_Option

Digital_Transfer_Information =
Format_Name +
([Format_Version_Number |
Format_Version_Date] +
(Format_Specification)) +
(Format_Information_Content) +
0{File_Decompression_Technique}1 +
(Transfer_Size)

Digital_Transfer_Option =
1{ [Online_Option |
Offline_Option] }n

Online_Option =
1{Computer_Contact_Information}n +
(Access_Instructions) +
(Online_Computer_and_Operating_System)

Computer_Contact_Information =
[Network_Address |
Dialup_Instructions]

Network_Address =
1{Network_Resource_Name}n

Dialup_Instructions =
Lowest_BPS +
0{Highest_BPS}1 +
Number_DataBits +
Number_StopBits +
Parity +
0{Compression_Support}1 +
1{Dialup_Telephone}n +
1{Dialup_File_Name}n

Offline_Option =
Offline_Media +
0{Recording_Capacity}1
1{Recording_Format}n +
0{Compatibility_Information}1

Recording_Capacity =
1{Recording Density}n +
Recording_Density_Units

3.1 **Distributor -- the party from whom the *tool for biological analysis* may be obtained.**
Type: compound

3.2 ***Local Identifier* -- the identifier by which the distributor knows the *tool for biological analysis. This field is applicable whenever there is an internal name or code by which the distributor will identify the tool for biological analysis.***
Type: text
Domain: free text

3.3 **Distribution Liability -- statement of the liability assumed by the**

distributor.

Type: text

Domain: free text

3.4 **Standard Order Process -- the common ways in which the *tool for biological analysis* may be obtained or received, and related instructions and fee information. *This field is not applicable if there is no standard method for obtaining the tool for biological analysis.***

Type: compound

3.4.1 **Non-digital Form -- the description of options for obtaining the *tool for biological analysis* on non-computer-compatible media.**

Type: text

Domain: free text

3.4.2 **Digital Form -- the description of options for obtaining the *tool for biological analysis* on computer-compatible media. *This may include executables, source code, sample data, and/or necessary files.***

Type: compound

3.4.2.1 **Digital Transfer Information - description of the form of the *information (this may include executables, source code, sample data, and/or necessary files)* to be distributed.**

Type: compound

3.4.2.1.1 **Format Name -- the name of the data transfer format *for executables, source code, sample data, and/or necessary files*.**

Type: text

Domain: domain values from the table below; free text

Domain Value	Definition
"ACC"	Access data base file
"ARCE"	ARC/INFO Export format
"ARCG"	ARC/INFO Generate format
"ASCII"	ASCII file, formatted for text attributes, declared format
"BIL"	Imagery, band interleaved by line
"BIP"	Imagery, band interleaved by pixel
"BSQ"	Imagery, band interleaved sequential
"EXCEL"	Excel data file
"CDF"	Common Data Format

"CFF" **Cartographic Feature File (U.S. Forest Service)**

"COORD" **User-created coordinate file, declared format**

"DBF" ***dBASE data file***

"DEM" **Digital Elevation Model format (U.S. Geological Survey)**

"DFAD" **Digital Feature Analysis Data (Defense Mapping Agency)**

"DGN" **Microstation format (Intergraph Corporation)**

"DIF" ***VisiCalc format***

"DIGEST" **Digital Geographic Information Exchange Standard**

"DLG" **Digital Line Graph (U.S. Geological Survey)**

"DOC" **Microsoft Word file**

"DTED" **Digital Terrain Elevation Data (MIL-D-89020)**

"DWG" **AutoCAD Drawing format**

"DX90" **Data Exchange '90**

"DXF" **AutoCAD Drawing Exchange Format**

"EPS" **Encapsulated Postscript**

"ERDAS" **ERDAS image files (ERDAS Corporation)**

"FW" ***Framework spreadsheet or database format***

"GIF" ***Graphics Interchange Format***

"GRA" **ARC/INFO graphic file**

"GRASS" **Geographic Resources Analysis Support System**

"HDF" **Hierarchical Data Format**

"IGDS" **Interactive Graphic Design System format (Intergraph Corporation)**

"IGES" **Initial Graphics Exchange Standard**

"MOSS" **Multiple Overlay Statistical System export file**

"netCDF" **network Common Data Format**

"NITF" **National Imagery Transfer Format**

"PBM" ***Portable Bit Map format file***

"PLT" ***ARC/INFO Plot file***

"PS" ***Postscript***

"QP" ***Quattro Pro data file***

"RPD" ***RapidFile***

"RPF" **Raster Product Format (Defense**

Mapping Agency)

"RVC" **Raster Vector Converted format (MicroImages)**

"RVF" **Raster Vector Format (MicroImages)**

"SDTS" **Spatial Data Transfer Standard (Federal Information Processing Standard 173)**

"SIF" **Standard Interchange Format (DOD Project 2851)**

"SLF" **Standard Linear Format (Defense Mapping Agency)**

"SPLUS" **S-Plus file**

"TIFF" **Tagged Image File Format**

"TGRLN" **Topologically Integrated Geographic Encoding and Referencing (TIGER) Line format (Bureau of the Census)**

"VPF" **Vector Product Format (Defense Mapping Agency)**

"WK1" ***Lotus 1-2-3 file***

"WKS" ***Lotus 1-2-3 file (older version than wk1)***

"WP" ***WorkPerfect***

3.4.2.1.2 **Format Version Number -- version number of the format.**

Type: **text**

Domain: **free text**

3.4.2.1.3 **Format Version Date -- date of the version of the format.**

Type: **date**

Domain: **free date**

3.4.2.1.4 **Format Specification -- name of a subset, profile, or product specification of the format.**

Type: **text**

Domain: **free text**

3.4.2.1.5 **Format Information Content -- description of the content of the data encoded in a format.**

Type: **text**

Domain: **free text**

3.4.2.1.6 **File Decompression Technique -- recommendations of algorithms or processes (including means of obtaining**

these algorithms or processes) that can be applied to read or expand *tool for biological analysis*s to which data compression techniques have been applied. *This field is applicable whenever a data file has been compressed. If some files are compressed but others are not, "No compression applied" would be applicable for the appropriate files. If compression is never used, this field is not applicable.*

Type: text

Domain: "No compression applied" free text

3.4.2.1.7 **Transfer Size -- the size, or estimated size, of the transferred *tool for biological analysis* in megabytes.**

Type: real

Domain: Transfer Size > 0.0

3.4.2.2 **Digital Transfer Option -- the means and media by which a *tool for biological analysis* is obtained from the distributor.**

Type: compound

3.4.2.2.1 **Online Option -- information required to directly obtain the *tool for biological analysis* electronically.**

Type: compound

3.4.2.2.1.1 **Computer Contact Information -- instructions for establishing communications with the distribution computer.**

Type: compound

3.4.2.2.1.1.1 **Network Address -- the electronic address from which the *tool for biological analysis* can be obtained from the distribution computer.**

Type: compound

3.4.2.2.1.1.1.1 **Network Resource Name -- the name of the file and service from which the *tool for biological analysis* can be obtained. (See description of "Network Addresses and File Names" on page xi.)**

Type: text

Domain: free text

3.4.2.2.1.1.2 **Dialup Instructions -- information required to access the distribution computer remotely through telephone lines.**

Type: compound

3.4.2.2.1.1.2.1 **Lowest BPS -- lowest or only speed for the connection's communication, expressed in bits per second.**
Type: integer
Domain: Lowest BPS >= 110

3.4.2.2.1.1.2.2 **Highest BPS -- highest speed for the connection's communication, expressed in bits per second. *This field is applicable* in cases when a range of rates are provided.**
Type: integer
Domain: Highest BPS > Lowest BPS

3.4.2.2.1.1.2.3 **Number DataBits -- number of data bits in each character exchanged in the communication.**
Type: integer
Domain: 7 <= Number DataBits <= 8

3.4.2.2.1.1.2.4 **Number StopBits -- number of stop bits in each character exchanged in the communication.**
Type: integer
Domain: 1 <= Number StopBits <= 2

3.4.2.2.1.1.2.5 **Parity -- parity error checking used in each character exchanged in the communication.**
Type: text
Domain: "None" "Odd" "Even" "Mark" "Space"

3.4.2.2.1.1.2.6 **Compression Support -- data compression available through the modem service to speed data transfer. *This field is applicable whenever data compression is available.***
Type: text
Domain: "V.32" "V.32bis" "V.42" "V.42bis" free text

3.4.2.2.1.1.2.7 **Dialup Telephone -- the telephone number**

of the distribution computer.
Type: text
Domain: free text

3.4.2.2.1.1.2.8 **Dialup File Name -- the name of a file containing the *tool for biological analysis* on the distribution computer.**
Type: text
Domain: free text

3.4.2.2.1.2 **Access Instructions -- instructions on the steps required to access the *tool for biological analysis*.**
Type: text
Domain: free text

3.4.2.2.1.3 **Online Computer and Operating System -- the brand of distribution computer and its operating system.**
Type: text
Domain: free text

3.4.2.2.2 **Offline Option -- information about media-specific options for receiving the *tool for biological analysis*.**
Type: compound

3.4.2.2.2.1 **Offline Media -- name of the media on which the *tool for biological analysis* can be received.**
Type: text
Domain: "CD-ROM" "3-1/2 inch floppy disk" "5-1/4 inch floppy disk" "9-track tape" "4 mm cartridge tape" "8 mm cartridge tape" "1/4-inch cartridge tape" free text

3.4.2.2.2.2 **Recording Capacity -- the density of information to which data are written. *This field is applicable whenever* different recording capacities are possible.**
Type: compound

3.4.2.2.2.2.1 **Recording Density -- the density in which the *tool for biological analysis* can be recorded.**
Type: real
Domain: Recording Density > 0.0

3.4.2.2.2.2.2 **Recording Density Units -- the units of measure for the recording density.**

Type: text
Domain: free text

3.4.2.2.2.3 **Recording Format -- the options available or method used to write the *tool for biological analysis* to the medium.**

Type: text
Domain: "cpio" "tar" "High Sierra" "ISO 9660" "ISO 9660 with Rock Ridge extensions" "ISO 9660 with Apple HFS extensions" free text

3.4.2.2.2.4 **Compatibility Information --- description of other limitations or requirements for using the medium. *This field is applicable whenever there are limitations or requirements for using the medium.***

Type: text
Domain: free text

3.4.3 **Fees -- the fees and terms for retrieving the *tool for biological analysis*.**

Type: text
Domain: free text

3.4.4 **Ordering Instructions -- general instructions and advice about, and special terms and services provided for, the *tool for biological analysis* by the distributor.**

Type: text
Domain: free text

3.4.5 **Turnaround -- typical turnaround time for the filling of an order.**

Type: text
Domain: free text

3.5 **Custom Order Process -- description of custom distribution services available, and the terms and conditions for obtaining these services. *This field is applicable whenever there is the possibility for custom ordering the tool for biological analysis.***

Type: text
Domain: free text

3.6 **Technical Prerequisites -- description of any technical capabilities that the consumer must have to use the *tool for biological analysis* in the form(s) provided by the distributor.**

Type: text

Domain: free text

3.7 Available Time Period -- the time period when the *tool for biological analysis* will be available from the distributor.

Type: date

Domain: free date

Metadata Reference Information

4 Metadata Reference Information -- information on the currentness of the metadata information, and the responsible party.

Type: compound

Metadata_Reference_Information =
Metadata_Date +
(Metadata_Review_Date +
(Metadata_Future_Review_Date)) +
Metadata_Contact +
Metadata_Standard_Name +
Metadata_Standard_Version +
(Metadata_Access_Constraints) +
(Metadata_Use_Constraints) +
(Metadata_Security_Information) +
(Metadata_Language)

Metadata_Contact =
Contact_Information ***(see section 6 for production rules)***

Metadata_Security_Information =
Metadata_Security_Classification_System +
Metadata_Security_Classification +
Metadata_Security_Handling_Description

4.1 Metadata Date -- the date that the metadata were created or last updated.

Type: date
Domain: free date

4.2 Metadata Review Date -- the date of the latest review of the metadata entry.

Type: date
Domain: free date; Metadata Review Date later than Metadata Date

4.3 Metadata Future Review Date -- the date by which the metadata entry should be reviewed.

Type: date
Domain: free date; Metadata Future Review Date later than Metadata Review Date

4.4 Metadata Contact -- the party responsible for the metadata information.

Type: compound

4.5 **Metadata Standard Name -- the name of the metadata standard used to document the *tool for biological analysis*.**

Type: text

Domain: "FGDC Content Standards for Digital Geospatial Metadata" *"NBS Content Standards for National Biological Information Infrastructure Metadata"* free text

4.6 **Metadata Standard Version -- identification of the version of the metadata standard used to document the *tool for biological analysis*.**

Type: text

Domain: free text

4.7 **Metadata Access Constraints -- restrictions and legal prerequisites for accessing the metadata. These include any access constraints applied to assure the protection of privacy or intellectual property, and any special restrictions or limitations on obtaining the metadata.**

Type: text

Domain: free text

4.8 **Metadata Use Constraints -- restrictions and legal prerequisites for using the metadata after access is granted. These include any access constraints applied to assure the protection of privacy or intellectual property, and any special restrictions or limitations on obtaining the metadata.**

Type: text

Domain: free text

4.9 **Metadata Security Information -- handling restrictions imposed on the metadata because of national security, privacy, or other concerns.**

Type: compound

4.9.1 **Metadata Security Classification System -- name of the classification system for the metadata.**

Type: text

Domain: free text

4.9.2 **Metadata Security Classification -- name of the handling restrictions on the metadata.**

Type: text

Domain: "Top secret" "Secret" "Confidential" "Restricted" "Unclassified" "Sensitive" free text

4.9.3 **Metadata Security Handling Description -- additional information about the restrictions on handling the metadata.**

Type: text

Domain: free text

4.10 ***Metadata Language -- Language(s) used within the metadata.***

Type: ***text***

Domain: ***free text***

Citation Information

5 **Citation Information -- the recommended reference to be used for the *tool for biological analysis*. *(Note: this section provides a means of stating the citation of a tool for biological analysis, and is used by other sections of the metadata standard. This section is never used alone.)***

Type: compound

Citation_Information =
1{Originator}n +
***Release_Date* +**
Title +
***(Short_Name)* +**
0{*Version*}1 +
0{*Form*}1 +
0{Series_Information}1 +
0{Publication_Information}1 +
0{Other_Citation_Details}1 +
(1{Online_Linkage}n)

Series_Information =
Series_Name +
Issue_Identification

Publication_Information =
Publication_Place +
Publisher

5.1 **Originator -- the name of an organization or individual that developed the *tool for biological analysis*. If the name of editors or compilers are provided, the name must be followed by "(ed.)" or "(comp.)" respectively. *If possible, the organization name should be given to the sub-organizational level to which the individual(s) that developed the data is (are) "attached".***

Type: text
Domain: "Unknown" free text

5.2 ***Release Date* -- the date when the *tool for biological analysis* is published or otherwise made available for release.**

Type: date
Domain: "Unknown" "Unpublished material" free date

5.3 **Title -- the *full* name by which the *tool for biological analysis* is known.**

Type: text

Domain: free text

5.4 ***Short Name -- the short name by which the tool for biological analysis is known. For example, the software package "Geographic Resources Analysis Support System" has a short name of "GRASS".***

Type: text
Domain: free text

5.5 ***Version* -- the version of the title.** ***This field is applicable whenever the tool for biological analysis has been released in multiple versions (i.e. updated).***

Type: text
Domain: free text

5.6 ***Form* -- the mode in which the *tool* is represented.** ***This field is applicable if the tool being cited is in a describable form.***

Type: text
Domain: *"C program" "dos executable" "java applet" "unix script" free text*

5.7 **Series Information -- the identification of the series publication of which the *tool for biological analysis* is a part.** ***This field is applicable whenever the tool for biological analysis is a subset of a larger number of related tool for biological analysiss, usually issued periodically over time.***

Type: compound

5.7.1 **Series Name -- the name of the series publication of which the *tool for biological analysis* is a part.**

Type: text
Domain: free text

5.7.2 **Issue Identification -- information identifying the issue of the series publication of which the *tool for biological analysis* is a part.**

Type: text
Domain: free text

5.8 **Publication Information -- publication details for published *tool for biological analysis*s.** ***This field is applicable whenever the information being cited has an identifiable publisher.***

Type: compound

5.8.1 **Publication Place -- the name of the city (and state or province, and country, if needed to identify the city) where the *tool for biological analysis* was published or released.**

Type: text
Domain: free text

5.8.2 **Publisher -- the name of the individual or organization that published the *tool for biological analysis*.**

Type: text

Domain: free text

5.9 **Other Citation Details -- other information required to complete the citation. *This field is applicable whenever there is additional information needed to adequately identify the cited material.***

Type: text

Domain: free text

5.10 **Online Linkage -- the name of an online computer resource that contains the *tool for biological analysis*. Entries should follow the Uniform Resource Locator convention of the Internet.**

Type: text

Domain: free text

Contact Information

6 **Contact Information -- Identity of, and means to communicate with, person(s) and organization(s) associated with the *tool for biological analysis*. *(Note: this section provides a means of identifying individuals and organizations, and is used by other sections of the metadata standard. This section is never used alone.)***

Type: compound

Contact_Information =
[Contact_Person_Primary |
Contact_Organization_Primary] +
(Contact_Position) +
1{Contact_Address}n +
1{Contact_Voice_Telephone}n +
(1{Contact_TDD/TTY_Telephone}n) +
(1{Contact_Facsimile_Telephone}n) +
(1{*Internet_Address*}n) +
(Hours_of_Service) +
(Contact_Instructions)

Contact_Person_Primary =
Contact_Person +
(Contact_Organization)

Contact_Organization_Primary =
Contact_Organization +
(Contact_Person)

Contact_Address =
Address_Type +
0{Address}n +
City +
State_or_Province +
Postal_Code +
(Country)

6.1 **Contact Person Primary -- the person, and the affiliation of the person, associated with the *tool for biological analysis*. Used in cases where the association of the person to the *tool for biological analysis* is more significant than the association of the organization to the *tool for biological analysis*.**

Type: compound

6.1.1 **Contact Person -- the name of the individual to which the contact type applies.**

Type: text
Domain: free text

6.1.2 **Contact Organization -- the name of the organization to which the contact type applies.**
Type: text
Domain: free text

6.2 **Contact Organization Primary -- the organization, and the member of the organization, associated with the *tool for biological analysis*. Used in cases where the association of the organization to the *tool for biological analysis* is more significant than the association of the person to the *tool for biological analysis*.**
Type: compound

6.3 **Contact Position -- the title of individual.**
Type: text
Domain: free text

6.4 **Contact Address -- the address for the organization or individual.**
Type: compound

6.4.1 **Address Type -- the information provided by the address.**
Type: text
Domain: "mailing address" "physical address" "mailing and physical address"

6.4.2 **Address -- an address line for the address. *This field is appropriate whenever there is a street address or PO Box number or similar address information for the contact.***
Type: text
Domain: free text

6.4.3 **City -- the city of the address.**
Type: text
Domain: free text

6.4.4 **State or Province -- the state or province of the address.**
Type: text
Domain: free text

6.4.5 **Postal Code -- the ZIP or other postal code of the address.**
Type: text
Domain: free text

6.4.6 **Country -- the country of the address.**
Type: text
Domain: free text

6.5 **Contact Voice Telephone -- the telephone number by which individuals can speak to the organization or individual.**
Type: text
Domain: free text

6.6 **Contact TDD/TTY Telephone -- the telephone number by which hearing-impaired individuals can contact the organization or individual.**
Type: text
Domain: free text

6.7 **Contact Facsimile Telephone -- the telephone number of a facsimile machine of the organization or individual.**
Type: text
Domain: free text

6.8 ***Internet Address*** **-- the address of the electronic mailbox of the organization or individual.**
Type: text
Domain: free text

6.9 **Hours of Service -- time period when individuals can speak to the organization or individual.**
Type: text
Domain: free text

6.10 **Contact Instructions -- supplemental instructions on how or when to contact the individual or organization.**
Type: text
Domain: free text

Appendix A

Glossary

[Most of the terms and definitions are from Department of Commerce, 1992, Spatial Data Transfer Standard (SDTS) (Federal Information Processing Standard 173): Washington: Department of Commerce, National Institute of Standards and Technology.]

tool for biological analysis -- a process or method that helps with the analysis, interpretation, application, or understanding of biological data. Tools help achieve results or create outcomes by helping people use data to solve problems. Examples of tools include ecological, biological, and physiological models, algorithms, integrated decisionsupport systems, simulations, and visualizations.

clearinghouse -- see National Geospatial Data Clearinghouse.

compound element -- a group of data elements and other compound elements. Compound elements represent higher-level concepts that cannot be represented by individual data elements.

data element -- a logically primitive item of data.

data set -- a collection of related data.

domain -- in the definition of the elements in the metadata standard, the domain identifies valid values for a data element.

geospatial data -- information that identifies the geographic location and characteristics of natural or constructed features and boundaries on the earth. This information may be derived from, among other things, remote sensing, mapping, and surveying technologies.

graph -- a set of topologically interrelated zero-dimensional (node), one-dimensional (link or chain), and sometimes two-dimensional (GT-polygon) objects that conform to a set of defined constraint rules. Numerous rule sets can be used to distinguish different types of graphs. Three such types, planar graph, network, and two-dimensional manifold, are used in this standard. All three share the following rules: each link or chain is bounded by an ordered pair of nodes, not necessarily distinct; a node may bound one or more links or chains; and links or chains may only intersect at nodes. Planar graphs and networks are two specialized types of graphs, and a two-dimensional manifold is an even more specific type of planar graph.

map -- a spatial representation, usually graphic on a flat surface, of spatial phenomena.

media -- the physical devices used to record, store, and (or) transmit data.

metadata -- data about the content, quality, condition, and other characteristics of data.
National Geospatial Data Clearinghouse -- a distributed network of geospatial data producers, managers, and users linked electronically. Building on initiatives such as the national information infrastructure, the clearinghouse uses a distributed, electronically connected network, such as the Internet. Each data provider will describe available data in an electronic form, and provide these descriptions (or "metadata") using means that can be accessed over a communications network. Thus, the data for the clearinghouse are located at the sites of data producers (or, where more efficient, at the sites of intermediaries) throughout the country. Using the network, users will search these descriptions to locate data that are suitable for their applications.

pixel -- two-dimensional picture element that is the smallest nondivisible element of a digital image.

primitive -- the quality of not being subdivided; atomic.

quality -- an essential or distinguishing characteristic necessary for data to be fit for use.

raster -- one or more overlapping layers for the same grid or digital image.

SDTS -- the Spatial Data Transfer Standard defined by Department of Commerce, 1992, Spatial Data Transfer Standard (SDTS) (Federal Information Processing Standard 173): Washington, Department of Commerce, National Institute of Standards and Technology.

spatial data -- see geospatial data.

type -- in the definition of the elements in the metadata standard, a compound element has the type "compound" to provide a unique way to identify compound elements. For a data element, the type identifies the kind of value that can be assigned to the data element. The choices are "integer" for integer numbers, "real" for real numbers, "text" for ASCII characters, "date" for day of the year, and "time" for time of the day.

Appendix B

Alphabetical List of Compound Elements and Data Elements

Appendix C

References

Federal Geographic Data Committee, 1994, Content Standards for Digital Geospatial Metadata: Reston, Virginia, Federal Geographic Data Committee.

National Biological Service, 1995, Draft Content Standard for National Biological Information Infrastructure Metadata: Denver, Colorado, National Biological Service.

American Congress on Surveying and Mapping and American Society of Civil Engineering, 1978, Definitions of surveying and associated terms: Falls Church, Virginia, American Congress on Surveying and Mapping.

American National Standards Institute, 1975, Representations of universal time, local time differentials, and United States time zone reference for information interchange (ANSI X3.51-1975): New York, American National Standards Institute.

American National Standards Institute, 1986, Representation for calendar date and ordinal date for information interchange (ANSI X3.30-1985): New York, American National Standards Institute.

American National Standards Institute, 1986, Representations of local time of day for information interchange (ANSI X3.43-1986): New York, American National Standards Institute.

American National Standards Institute, 1990, Dictionary for information systems (ANSI X3.172-1990): New York, American National Standards Institute.

Anglo-American Committee on Cataloguing of Cartographic Materials, 1982, Cartographic materials: A manual of interpretation for AACR2: Chicago, American Library Association.

ASTM Section D18.01.05, various dates, Spatial metadata content standards for geographic information systems, catalogs, and data exchange (drafts).

Clark, Suzanne, Larsgaard, Mary, and Teague, Cynthia, 1992, Cartographic citations: A style guide: Chicago, American Library Association, Map and Geography Roundtable.

Cogan, Christopher, and Edwards, Thomas, Jr., 1994 (February), Metadata standards for Gap analysis: Moscow, Idaho, Idaho Cooperative Fish and Wildlife Research Unit, University of Idaho.

Department of Commerce, 1986, Representation of geographic point locations for information interchange (Federal Information Processing Standard 70-1): Washington, Department of Commerce, National Institute of Standards and Technology.

Department of Commerce, 1989 (January), State Plane Coordinate System of 1983 (National Oceanic and Atmospheric Administration Manual NOS NGS 5): Silver Spring, Maryland, Department of Commerce, National Oceanic and Atmospheric Administration, National Ocean Service, Coast and Geodetic Survey.

Department of Commerce, 1992, Spatial Data Transfer Standard (SDTS) (Federal Information Processing Standard 173): Washington, Department of Commerce, National Institute of Standards and Technology.

Department of Defense, 1981, Glossary of mapping, charting, and geodetic terms (4th ed.): Washington, Department of Defense, Defense Mapping Agency.

Department of Defense, 1990, Military specification ARC Digitized Raster Graphics (ADRG) (MIL-A-89007): Philadelphia, Department of Defense, Defense Printing Service Detachment Office.

Department of Defense, 1992, Vector Product Format (MIL-STD-600006): Philadelphia, Department of Defense, Defense Printing Service Detachment Office.

Dodd, Susan, 1982, Cataloging machine-readable data files: Chicago, American Library Association.

Hansen, Wallace, 1991, Suggestions to authors of the reports of the United States Geological Survey (7th ed.): Washington, U.S. Government Printing Office.

Li, Xia, and Crane, Nancy, 1993, Electronic style: A guide to citing electronic information: Westport, Connecticut, Meckler Publishing.

Network Development and MARC Standards Office, 1988, USMARC format for bibliographic data: Washington, Library of Congress, Cataloging Distribution Service.

Network Development and MARC Standards Office, 1988, USMARC code list for relators, sources, and description conventions: Washington, Library of Congress, Cataloging Distribution Service.

(no author), 1994, The Government Information Locator Service (GILS): Report to the Information Infrastructure Task Force (May 2, 1994).

Patrias, Karen, 1991 (April), National Library of Medicine recommended formats for bibliographic citations: Bethesda, Maryland, U.S. Department of Health and Human Services,

Public Health Service, National Institutes of Health, National Library of Medicine.

***Snyder*, John, 1987, Map projections: A working manual (U.S. Geological Survey Professional Paper 1395): Washington, U.S. Government Printing Office.**

Westbrook, J. H., and Grattidge, W., 1991, A glossary of terms relating to data, data capture, data manipulation, and data bases: CODATA Bulletin, v. 23, no. 1-2.